Article Watch: December 2018

Clive A. Slaughter

This column highlights recently published articles that are of interest to the readership of this publication. We encourage ABRF members to forward information on articles they feel are important and useful to Clive Slaughter, MCG-UGA Medical Partnership, 1425 Prince Ave., Athens, GA 30606, USA. Tel: (706) 713-2216; Fax: (706) 713-2221; E-mail: cslaught@uga.edu, or to any member of the editorial board. Article summaries reflect the reviewer's opinions and not necessarily those of the association.

NUCLEIC ACID SEQUENCING

Adiconis X, Haber A L, Simmons S K, Levy Moonshine A, Ji Z, Busby M A, Shi X, Jacques J, Lancaster M A, Pan J Q, Regev A, Levin J Z. Comprehensive comparative analysis of 5′-end RNA-sequencing methods. *Nature Methods* 15; 2018:505–511.

In the human genome, 54% of genes are annotated to have multiple transcription start sites. The use of alternate transcription start sites may affect gene function both in normal physiology and in disease. Adiconis *et al.* compare 6 published protocols by which RNAseq may be used to identify the 5′ ends of transcripts: cap analysis of gene expression (CAGE), RNA annotation and mapping of promoters for the analysis of gene expression (RAMPAGE), single-cell tagged reverse transcription (STRT), nano-cap analysis of gene expression (nanoCAGE), and oligo-capping, and global nuclear run-on cap (GRO-cap). The methods vary in the amount of RNA they require: STRT, which was developed for low-input amounts, requires the least RNA, whereas oligo-capping requires the most. The lowest recommended amount of RNA for each method was used in the comparison. CAGE is identified as the overall best method for identifying the 5′ end of mRNAs, although even this method yields an appreciable fraction of reads that map to regions far from the 5′ end (24%). Variation in specificity among the methods is replicated when testing spike-in RNAs derived from the External RNA Controls Consortium (ERCC). Interestingly, differences in 5′ specificity were noted among individual spike-in transcripts.

Giraldez M D, Spengler R M, Etheridge A, Godoy P M, Barczak A J, Srinivasan S, De Hoff P L, Tanriverdi K, Courtright A, Lu S, Khoory J, Rubio R, Baxter D, Driedonks T A P, Buermans H P J, Nolte-'t Hoen E N M, Jiang H, Wang K, Ghiran I, Wang Y E, Van Keuren-Jensen K, Freedman J E, Woodruff P G, Laurent L C, Erle D J, Galas D J, Tewari M. Comprehensive multicenter assessment of small RNA-seq methods for quantitative miRNA profiling. *Nature Biotechnology* 36;2018:746.

The nature and extent of technical biases introduced in the process of cDNA synthesis when mRNA expression is studied by RNA-seq have been well characterized. When RNA-seq is used to study small RNAs, however, the methods for cDNA synthesis are different. The biases associated with them have been less well characterized. In the present report, 9 laboratories from the NIH-funded Extracellular Communication Consortium assess different library preparation protocols for accuracy, reproducibility, and technical bias in the results of small RNA-seq. They use standardized, synthetic reference reagents and reference RNA samples of biologic origin. With the analysis of equimolar mixtures of small RNAs, the authors show that sequence-related bias is very large, reaching a factor of 10^4 with some of the commonly used commercial library preparation protocols. Library preparation protocols that use adaptors with degenerate bases in the ligating ends reduce the magnitude of this bias by a factor of 100, but the residual bias remains large and precludes the use of read counts alone for accurate quantification of different small RNAs in a sample. The level of bias varies among protocols even when degenerate adaptors are used. This indicates that additional variables affect outcome. Nevertheless, the study indicates that small RNA-seq provides consistently accurate relative quantification of a given miRNA among samples, provided the same library preparation method is used. The samples of biologic origin showed consistent profiling among laboratories when the same preparation protocol is used. RNA-seq is increasingly used for study of materials with very low RNA concentrations, including the extracellular vesicles and RNA protein complexes in biofluids. Standardization and normalization methods become increasingly important in the study of such materials. The present work provides

doi: 10.7171/jbt.18-2904-001

benchmarking criteria for the normalization of small RNA-seq datasets and alerts investigators to the potential for inaccuracy resulting from limitations in sequencing coverage that might be introduced with methods that confer higher levels of bias.

Lentini A, Lagerwall C, Vikingsson S, Mjoseng H K, Douvlataniotis K, Vogt H, Green H, Meehan R R, Benson M, Nestor C E. A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nature Methods* 15;2018:499–504.

Patterns of genomic methylation are often studied by methods involving enrichment of methylated DNA by immunoprecipitation. Anti-5-methylcytosine (5mC) and anti-5-hydroxymethylcytosine (5hmC) antibodies are commonly used, and high-throughput DNA sequencing is employed to identify regions of the genome enriched or depleted during the immunoprecipitation. This approach is referred to as DNA immunoprecipitation sequencing (DIP-seq). DNA before immunoprecipitation is often used as the control. The DIP-seq method has also been extended to the mapping of additional DNA modifications. The present paper documents various sources of systematic error in such experiments. The authors indicate that all of the commonly used antibodies, regardless of their nominal specificity, bind to short tandem repeat (STR) sequences regardless of DNA modification. Preimmune IgG binds similarly to STRs. The authors report that 95% of published DIP-seq studies do not use the appropriate preimmune control to identify such off-target effects and that 50–99% of the regions enriched in DIP-seq are false positives. Controlling for off-target effects improves the signal-to-noise ratio and strongly diminishes overlap between 5mC- and 5hmC-marked genes. The authors also investigate the occurrence of 6-methyladenosine (6mA), a modified base that is very rare in mammalian cells but abundant in cultures infected with *Mycoplasma*. Even low-level bacterial contamination yields sequencing reads that match genomic sequences of the suspected contaminants. The results of this study indicate that preimmune IgG controls should be used in DIP-seq experiments, and findings should be validated using nonantibody-based techniques to avoid misinterpretation.

GLYCANS

Remoroza C A, Mak T D, De Leoz M L A, Mirokhin Y A, Stein S E. Creating a mass spectral reference library for oligosaccharides in human milk. *Analytical Chemistry* 90;2018:8977–8988.

Human milk contains oligosaccharides composed of 3–10 monosaccharide units. The monosaccharides include glucose (Glc), galactose (Gal), *N*-acetyl glucosamine (GlcNAc), fucose (Fuc), and sialic acid (Neu5Ac). The oligosaccharides commonly have a lactose (Galβ1→4Glc) or *N*-acetyl-lactosamine (Galβ1→4GlcNAc) disaccharide sequence at the reducing end. Reference standards for many of the milk oligosaccharides are unavailable commercially. In this context, Remoroza *et al.* seek to facilitate mass spectral analysis of human milk by making available a spectral library of milk oligosaccharides to enable their identification in unknown samples. The authors perform mass spectrometry to identify oligosaccharides from the National Institute of Standards and Technology (NIST) Standard Reference Material (SRM) 1953, which consists of human milk pooled from 100 breastfeeding mothers. They use hydrophilic interaction liquid chromatography and electrospray ionization. Fragment ion mass spectra are acquired in an Orbitrap mass spectrometer by collisional dissociation (CD) or higher energy collisional dissociation (HCD). Structural assignment is performed by searching the NIST 17 Tandem MS Library with a search algorithm that permits matching despite m/z shifts. The authors assign 74 oligosaccharides in this way. These include novel glycan structures. The resulting SRM 1953 oligosaccharide mass spectral library is anticipated to help identify oligosaccharides both in milk and in other biologic fluids.

Duan J, Amster I J. An automated, high-throughput method for interpreting the tandem mass spectra of glycosaminoglycans. *Journal of The American Society for Mass Spectrometry* 29;2018:1802–1811.

Duan and Amster propose a new algorithm for automated interpretation of fragment ion mass spectra of glycosaminoglycans (GAGs). GAGs are a family of diverse molecules with a nontemplated structure for which neither an adequate set of reference standards nor a structure database is available. GAG structure is based on a simple repeating copolymer of glucuronic acid and *N*-acetyl galactosamine, but modifications (*O*-sulfation, *N*-deacetylation/sulfonation, and uronic acid stereochemistry) are dispersed along the chain in ways that influence the biologic properties of the polymer. The resulting fragment ion mass spectra are complex, and their interpretation is made even more difficult by the facile loss of sulfate modifications during fragmentation. The authors have developed an algorithm that lists possible fragment ion signals for each candidate structure and then compares the experimental signals against each theoretical fragment list and ranks the goodness-of-fit for each permutation. A genetic algorithm limits the number of permutations of isomeric structures to be considered and thereby controls search time. The method can be used with GAGs of indefinite length. The authors test it on fragment ion data from long-chain, moderately sulfated GAGs of the chondroitin sulfate family derived from the proteoglycan

bikunin. The automated method yields the same structures as those reported in the literature.

MACROMOLECULAR SYNTHESIS AND SYNTHETIC BIOLOGY

Palluk S, Arlow D H, de Rond T, Barthel S, Kang J S, Bector R, Baghdassarian H M, Truong A N, Kim P W, Singh A K, Hillson N J, Keasling J D. De novo DNA synthesis using polymerase-nucleotide conjugates. *Nature Biotechnology* 36;2018:645.

Phosphoramidite chemistry has long dominated the synthesis of oligonucleotides, but Palluk *et al.* report progress in developing an enzymatic method. They use a terminal deoxynucleotidyl transferase to add nucleotides sequentially without a template. Each enzyme molecule is covalently conjugated to a single nucleotide *via* a cleavable linker. The enzyme couples its cargo to the 3′ end of the growing chain, and once coupled, the bound enzyme molecule acts as a sterically bulky protecting moiety that prevents further chain extension. The enzyme is linked to its nucleotide *via* a photo-cleavable, amine-to-thiol linker. After each nucleotide addition, the enzyme is released by photolysis. A propargylamino group remains on the base, but the authors demonstrate that a natural complementary DNA strand may be accurately synthesized from a template of the *N*-acetyl-propargylamino oligonucleotide by Taq polymerase. The sequence of the assembled DNA is determined by the order of addition of enzyme-nucleotide conjugates. One residue can be added in 10–20 s. Oligonucleotides of up to 10 residues have been assembled in proof-of-principle experiments with stepwise yields of 93.4–99.5%. Further development of this methodology may lead to a viable alternative to standard chemical oligonucleotide synthesis.

Barber K W, Muir P, Szeligowski R V, Rogulina S, Gerstein M, Sampson J R, Isaacs F J, Rinehart J. Encoding human serine phosphopeptides in bacteria for proteome-wide identification of phosphorylation-dependent interactions. *Nature Biotechnology* 36;2018: 638.

Barber *et al.* build on methods recently developed for expressing phosphorylated proteins in *E. coli*, recoded for genetically programmed incorporation of the nonstandard amino acid phosphoserine (pSer). The investigators design oligonucleotides that encode >110,000 peptide sequences spanning known serine phosphorylation sites. They include 15 amino acids on either side of each pSer. They synthesize the oligonucleotide library on a programmable DNA microarray and include universal primer annealing and restriction sites. The components of the library are then assembled into an expression vector. The central pSer residue in each peptide sequence is encoded by a UAG codon. In response to this codon, bacteria incorporate either Ser or pSer, depending on the translation system supplied. The investigators proceed to use their expression system for high-throughput identification of phosphorylated sequences that interact with protein domains of interest. For this purpose, a phosphopeptide-encoding sequence fused with the N-terminal portion of a split mCherry reporter is incorporated into a plasmid that also encodes the protein domain of interest fused with the C-terminal portion of mCherry. When a phosphopeptide binds to the protein domain, the two split portions of mCherry can interact, thereby restoring mCherry fluorescence. Florescent cells in which this occurs are isolated by fluorescence-activated cell sorting, and the interacting phosphorylated peptides are identified by high-throughput DNA sequencing. Many new phosphopeptide sequences that interact with 14-3-3 proteins or WW domains are identified by the authors in this way and may now be investigated to reveal potentially new protein–protein interactions. The new methodology is anticipated to facilitate investigation of phosphoproteome function, and its use may be extended to investigation of other post-translational modifications.

Luo J, Sun X, Cormack B P, Boeke J D. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* 560;2018: 392–396.

Shao Y, Lu N, Wu Z, Cai C, Wang S, Zhang L-L, Zhou F, Xiao S, Liu L, Zeng X, Zheng H, Yang C, Zhao Z, Zhao G, Zhou J-Q, Xue X, Qin Z. Creating a functional single-chromosome yeast. *Nature* 560; 2018:331–335.

Two groups report experiments to reduce the normal number of chromosomes (16) in haploid cells of the yeast *Saccharomyces cerevisiae* to investigate the effects of karyotype on gene expression, cell viability, and reproductive isolation. With the use of CRISPR-Cas9 genome editing, both groups progressively decrease the haploid chromosome number by successive rounds of end-to-end chromosome fusion with elimination of intervening telomeric sequences at each stage. They accompany these changes with appropriate deletion of centromeres to endure the presence of only 1 centromere per chromosome. Luo *et al.* generate a strain with just 2 chromosomes, and Shao *et al.* succeed in producing a strain with just 1 chromosome. The strains remain otherwise isogenic with wild-type yeast. The new stains are phenotypically similar to wild type and have similar gene-expression patterns, although they display somewhat reduced growth and competitiveness. The capacity to form viable spores upon mating between strains with different haploid numbers of chromosomes, however, is strongly impaired. Shao *et al.* further show that the chromatin 3-D structure is affected by changing the chromosome number.

The ability to engineer yeast strains with arbitrary numbers of chromosomes opens new opportunities to test the effects of centromeric and telomeric sequences on gene function.

Dou J, Vorobieva A A, Sheffler W, Doyle L A, Park H, Bick M J, Mao B, Foight G W, Lee M Y, Gagnon L A, Carter L, Sankaran B, Ovchinnikov S, Marcos E, Huang P-S, Vaughan J C, Stoddard B L, Baker D. De novo design of a fluorescence-activating β-barrel. *Nature* 561;2018:485–491.

Significant progress is now being made in the *de novo* design of proteins to adopt certain predetermined folding patterns and in redesigning existing proteins to bind designated small molecules. Dou *et al.* announce a successful *de novo* design of proteins with an antiparallel β-barrel-folding motif and demonstrate full *de novo* design of a β-barrel that binds a particular small molecule. The authors determine by simulation that symmetrical parametric models with uniform geometry are unable to achieve regular hydrogen bonding within the β-barrel scaffold and yield structures with an unacceptable strain. Therefore, they insert glycine residues to introduce kinks in β-strands and in turns for relief of the strain. From 500 designs formulated in this way, they choose the 4 most favorable for synthesis and characterization. Two form presumptive β-barrel structures, one as a monomer and the other as a tetramer. X-Ray crystallography of the monomer confirms its structural concordance with the design model. The authors then build a family of backbones for which cavity shapes match the shape of the small molecule fluorophore DFHBI, which fluoresces when constrained in a planar conformation upon binding to a protein. They finally undertake searches for optimal amino acid side chains and ligand placement within the protein cavity to achieve complementarity with the ligand. In this way, they identify designs that bind and activate the fluorescence of DFHBI *in vitro* and in *E. coli*, yeast, and mammalian cells. This work represents a new milestone in the *de novo* design of functional proteins.

MASS SPECTROMETRY

Riley N M, Sikora J W, Seckler H S, Greer J B, Fellers R T, LeDuc R D, Westphall M S, Thomas P M, Kelleher N L, Coon J J. The value of activated ion electron transfer dissociation for high-throughput top-down characterization of intact proteins. *Analytical Chemistry* 90;2018:8553–8560.

Rush M J P, Riley N M, Westphall M S, Coon J J. Top-down characterization of proteins with intact disulfide bonds using activated-ion electron transfer dissociation. *Analytical Chemistry* 90;2018:8946–8953.

The research group authoring these 2 papers previously showed that the coupling of infrared photoactivation with electron transfer dissociation (ETD)—so-called activated ion ETD (AI-ETD)—helps to disrupt noncovalent attachments that persist during formation of fragment ions by ETD and thus, greatly improves mass spectral sequence coverage of small, intact proteins. Riley *et al.* now couple AI-ETD with online, reverse-phase chromatographic separation of intact proteins. They test the performance of this methodology against standard higher energy-collisional dissociation (HCD) in an application involving analysis of the proteoforms of intact proteins (<20 kDa) in a human colorectal cancer cell line. They identify nearly as many proteoforms by AI-ETD as by HCD but note improved spectral quality and sequence coverage with AI-ETD that permits localization of post-translational modifications with enhanced confidence. Rush *et al.* pursue the comparison between methodologies by characterizing proteins heavily bridged with intrachain disulfide bonds. With the use of standard proteins, they observe better performance of AI-ETD in sequence coverage, fragment ion generation, and disulfide cleavage and note a particularly strong advantage of AI-ETD over ETD for precursor ions of low-charge density, which are poorly fragmented by ETD.

Hebert A S, Prasad S, Belford M W, Bailey D J, McAlister G C, Abbatiello S E, Huguet R, Wouters E R, Dunyach J-J, Brademan D R, Westphall M S, Coon J J. Comprehensive single-shot proteomics with FAIMS on a hybrid Orbitrap mass spectrometer. *Analytical Chemistry* 90;2018:9529–9537.

Hebert *et al.* benchmark a new implementation of the field asymmetric ion mobility spectrometry (FAIMS) in conjunction with an Orbitrap hybrid mass spectrometer for use in proteomic analysis of peptide ions. FAIMS separates gas-phase ions according to their mobility in an electric field (the dispersion field), while the ions travel though an inert gas. Ions are transported by the carrier gas between 2 electrodes that impose the dispersion field in a direction orthogonal to the direction of flow. The field alternates between a high voltage for a short period of time and a low voltage of the opposite polarity for a longer period of time. Ions alternate between traveling toward 1 electrode and then toward the other as the field oscillates. The oscillation times may be set so that ions that behave with the same mobility under both polarities are transmitted, whereas ions that differ in mobility under the 2 polarities impinge on the electrodes and are lost. However, a direct current compensation voltage may be added to the waveform to bias the net drift to permit passage of different subsets of ions through the device. The

new device incorporates features that provide improved maximal ion transmission, including better ion sampling at the interface between the electrospray source and the FAIMS sector, and a smaller gap between the cylindrical electrodes, which allows a higher dispersion field strength and shorter ion transit time. With this device, the authors quantify 7818 proteins in a single-shot, 4-h analysis, during which 4 different compensation voltages are applied. Without FAIMS, 6809 proteins are quantified. The authors suggest that the performance of the new system may shift the balance between convenience of data acquisition without chromatographic prefractionation and the enhancement of coverage available with chromatography.

FUNCTIONAL GENOMICS AND PROTEOMICS

Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology* 36;2018:765.

Kosicki *et al.* have observed unexpected results of CRISPR-Cas9 gene editing in mitotically active cells. The new results indicate a need for caution in clinical application of the technology. In experiments to disrupt coding sequences of the X-linked *PigA* gene in mouse embryonic stem cells, the authors supply guide RNAs (gRNAs) targeted to exonic sites, and, in parallel control experiments, supply gRNAs targeted to intronic sites hundreds to thousands of base pairs distant, expecting that these would not cause gene disruption. Surprisingly, however, they observe that the negative controls yield PigA-deficient cells in 5–20% of instances, indicating the possibility that repair of on-target, double-stranded breaks introduced by Cas9, widely believed to be highly localized and specific in the repair process mediated by nonhomologous end-joining (insertion or deletion of $<$20 bp), might sometimes yield much more extensive deletions in the kilobase range. The investigators proceeded to demonstrate that these effects occur commonly at transcriptionally active genes in all cell lines tested and are not limited to particular methods of Cas9 expression or delivery. They confirm the presence of large deletions using long-range PCR and sequencing by the Sanger method and by the Pacific Biosystems platform. They further demonstrate complex lesions involving insertion of DNA sequences from other chromosomes, inversions, duplications, and noncontiguous lesions. Such genetic damage would be undetected by the short-range PCR assays commonly used to monitor the results of gene editing. The authors caution that mutations that juxtapose a target gene to a cancer-driver gene could initiate or predispose to neoplasia and urge that comprehensive genomic analysis be deployed in *ex vivo* gene-editing studies going forward.

Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine* 24;2018:927–930.

Ihry R J, Worringer K A, Salick M R, Frias E, Ho D, Theriault K, Kommineni S, Chen J, Sondey M, Ye C, Randhawa R, Kulkarni T, Yang Z, McAllister G, Russ C, Reece-Hoyes J, Forrester W, Hoffman G R, Dolmetsch R, Kaykas A. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nature Medicine* 24;2018:939–946.

Two groups identify and mitigate a limitation in the efficiency of genome editing by the CRISPR-Cas9 methodology. Haapaniemi *et al.*, working with immortalized human retinal pigment epithelial cells, and Ihry *et al.*, working with human pluripotent stem cells, observe that the double-stranded breaks in DNA, induced by Cas9, are cytotoxic and kill the majority of treated cells. Double-stranded breaks trigger a response involving cell-cycle arrest and programmed cell death that normally functions to protect against genomic damage. This process is mediated by the tumor-suppressor protein p53. Both groups demonstrate that editing efficiency is substantially improved by transient suppression of this p53-dependent cytotoxic response. These observations have important implications for genome editing in the clinical context. On the negative side, suppression of the p53-driven responses exposes cells to increased risk of unintended DNA damage during the gene-editing process. On the positive side, such suppression minimizes selective pressure that favors cells with a deficient p53 tumor-suppressor pathway that may arise during the gene-editing process. These considerations further underscore the importance of comprehensive genome analysis in characterizing cells subjected to genome editing by CRISPR-Cas9.

PROTEIN CHARACTERIZATION

Nogly P, Weinert T, James D, Carbajo S, Ozerov D, Furrer A, Gashi D, Borin V, Skopintsev P, Jaeger K, Nass K, Båth P, Bosman R, Koglin J, Seaberg M, Lane T, Kekilli D, Brünle S, Tanaka T, Wu W, Milne C, White T, Barty A, Weierstall U, Panneels V, Nango E, Iwata S, Hunter M, Schapiro I, Schertler G, Neutze R, Standfuss J. Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* 361;2018:eaat0094.

Bacteriorhodopsin is an archaeal integral membrane protein that transduces light energy to perform electrochemical

work. Its duty cycle is initiated by photochemical isomerization of the ligand retinal from the all-*trans* conformation to a *cis* configuration about the C_{13}=C_{14} bond. This transition requires concomitant, subtle changes in the conformation of the polypeptide. Spectroscopic measurements indicate that these peptide conformational changes occur on a picosecond time scale. The present paper describes the conformational changes that occur at the beginning of the photochemical process, as revealed by time-resolved serial femtosecond crystallography. This technique uses an X-ray free-electron laser. A light pulse from a visible laser initiates the reaction sequence in a bacteriorhodopsin microcrystal, and after a controlled time delay, the crystal is illuminated by pulse from an X-ray laser, and a diffraction pattern is acquired. The crystal is destroyed by the X-ray pulse. With the use of the coherent X-ray imaging beamline at the Linac Coherent Light Source at Stanford University, Nogly *et al.* collect ~1 million diffraction patterns at 1.50 Å resolution. They calculate overlapping difference Fourier electron-density maps at 4 sequential time intervals within the first picosecond following the X-ray pulse (time resolution ~200 fs), as well as further datasets at 10 ps and 8.33 ms. The data yield 20 snapshots of bacteriorhodopsin activation that define the trajectory of retinal isomerization and details of polypeptide side chains and functional water molecules at near-atomic resolution. The authors thus capture the structural transitions constituting one of the fastest processes in biology.

CELL BIOLOGY AND TISSUE ENGINEERING

Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, Mali P, Church G M. Developmental barcoding of whole mouse via homing CRISPR. *Science* 361; 2018:eaat9804.

In this study, the technique of *in vivo* barcoding for reconstruction of developmental cellular lineages is extended to mammalian systems. As in earlier applications of the barcoding methodology, when Cas9 is active, a guide RNA targets mutation at the same gene that encodes this guide RNA. The process allows multiple rounds of mutation at this gene and generates evolutionary sequence changes that are amenable to analysis by single-cell RNA sequencing, thereby permitting lineage relationships among the host cells to be delineated. The challenge of creating a barcoding system appropriate for the complexity of mammalian development is met in the present study by deployment of multiple barcoding elements (genes) that vary in sequence independently. Ten such genes are estimated to be capable of generating 10^{23} distinct theoretical barcode combinations. The barcoding process is initiated when a mouse line bearing 60 target genes with stable sequences in the absence of Cas9 is crossed with a Cas9 knock-in mouse line. The authors observe that different target genes have different activity profiles; some mutate soon after conception, whereas others mutate more slowly throughout gestation. This provides a favorable spread of barcoding throughout development. Among the results illustrating the capabilities of the system acquired in this study is one derived by sequencing barcodes from the left and right sides of the forebrain, midbrain, and hindbrain. The barcodes from the right and left sides of each region are more closely related than the barcodes from different regions. This indicates that anterior–posterior commitment precedes lateral commitment during embryogenesis.

Wang X, Allen W E, Wright M A, Sylwestrak E L, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J, Nolan G P, Bava F-A, Deisseroth K. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361;2018:eaat5691.

Wang *et al.* combine *in situ* single-cell RNA sequencing with hydrogel tissue clarification technology to perform expression analysis of more than a thousand genes while retaining the 3-dimensional cellular anatomy of a tissue. They name the technique spatially resolved transcript amplicon readout mapping (STARmap). It incorporates a process for *in situ* amplification of cDNA probes hybridized to cellular mRNAs that does not depend on reverse transcription, an inherently inefficient process. To each mRNA, a pair of target-specific probes is hybridized: a primer probe and a padlock probe. If—and only if—both probes are hybridized to the same mRNA molecule, then the padlock probe will be circularized and rolling circle amplification of the padlock probe initiated. This process maximizes signal intensity and signal-to-noise ratio. During this synthesis of DNA amplicons, a low concentration of 5-(3-aminoallyl)-dUTP is supplied, and this residue is incorporated at a low level in place of T. This residue is further functionalized with an acrylamide moiety using acrylic acid *N*-hydroxysuccinimide ester, and is copolymerized with acrylamide monomers to fix amplicons in place within the polymer matrix. Proteins are then digested and lipids removed to render the matrix transparent. A library of 1024, 5-base barcodes is incorporated in the padlock probes as gene-unique identifiers. For *in situ* readout of sequence abundance, these identifiers are detected using a special sequencing-by-ligation method. Two kinds of probe are used: a short (11-nt), partially degenerate reading probe to set the base position to be interrogated and an 8-nt fluorescent detector probe with color coding of the dinucleotide at its 3′ end to transduce the base information for fluorescent imaging. A T4 DNA ligase

ligates the 2 probes together when they are perfectly complementary to the target. The short-reading and decoding probes are washed away, leaving the 19-nt product stably hybridized to its perfectly matched amplicon for imaging. After imaging, the probes can be stripped away and annealed to detect another set of mRNAs. In this way, STARmap is shown to detect 1000 genes over 6 imaging cycles. Wang *et al.* apply this technology to study the mouse visual cortex. They cluster gene expression patterns into a dozen distinct neuronal and non-neuronal cell types, map their spatial distribution across layers of the cortex, and document upregulation of activity-regulated genes in response to visual stimulation. Turning toward the more complex prefrontal cortex, they then visualize short- and long-range spatial organization of the cortical neurons by measuring expression of 1020 genes in millimeter-scale volumes containing >30,000 cells. This study foreshadows new possibilities for the analysis of functional interconnections among the cells comprising complex organ systems.

Wong B G, Mancuso C P, Kiriakov S, Bashor C J, Khalil A S. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nature Biotechnology* 36;2018:614.

Experimentation to study microbial evolution or direct such evolution for selection of new, functional characteristics has long been a central feature of microbial science. The main challenge in such experimentation is the achievement of a balance between throughput and precision of control: bioreactors are highly controllable but low-throughput, whereas batch reactors are less amenable to stringent control but are higher in throughput. Wong *et al.* describe the architecture of a do-it-yourself, automated system that affords both throughput and control stringency within the same hardware framework by providing multiple, precisely defined growth or selection conditions simultaneously. The system is modular. Firstly, a mass-produced sleeve unit that accommodates a 40-mL reaction vessel is equipped with sensors, actuators, and other electronic components mounted on a printed circuit board. Large numbers of such sleeve units can be accommodated within the same system for control of the desired culture parameters. Secondly, a fluidic module controls movement of media, culture, and reagents within the system. This module may be configured either with peristaltic pumps or with a complex, millifluidic system that incorporates principles of large-scale integration adopted for electronic and microfluidic devices. Thirdly, an electronics module communicates with the sleeve units *via* a motherboard and may be controlled remotely *via* the web. This system is tested in studies to evolve yeast populations under multiple selection conditions, to perform growth selection for yeast knockout libraries under temporally variable stress conditions, and to execute complex fluidic manipulations. The authors envisage extension of applications for the system. The contributions of individual species to community fitness in microbial consortia may be studied in this way, and circuits for producer strains may be designed to maximize temporal stability in industrial bioreactors.

DRUG DISCOVERY, DESIGN, AND MANUFACTURE

Isabella V M, Ha B N, Castillo M J, Lubkowicz D J, Rowe S E, Millet Y A, Anderson C L, Li N, Fisher A B, West K A, Reeder P J, Momin M M, Bergeron C G, Guilmain S E, Miller P F, Kurtz C B, Falb D. Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nature Biotechnology* 36;2018:857.

The authors of this report use synthetic biology and microbiome research to develop live bacterial therapeutics for treatment of inborn errors of metabolism. They have engineered *Escherichia coli* Nissle, a strain that does not colonize humans, to express pathways that degrade phenylalanine, and are testing it for efficacy in the management of phenylketonuria (PKU). The strain combines phenylalanine ammonia lyase with a high-affinity phenylalanine transporter to convert phenylalanine to *trans*-cinnamate. The genes are placed under the control of an anaerobic-inducible promoter to respond to the anoxic conditions in the mammalian gastrointestinal tract. The strain additionally incorporates the membrane-associated L-amino acid deaminase to convert phenylalanine to phenylpyruvate. This enzyme is oxygen dependent and is included to make use of the available oxygen in the proximal GI tract at the time of dosing. When the engineered bacteria are administered to a PKU mouse model, blood concentrations of phenylalanine decrease by 38% compared with controls, independent of dietary protein intake. In healthy cynomolgus monkeys (no primate PKU model is available), the bacteria inhibit increases in the serum phenylalanine level after oral phenylalanine challenge. The host converts the *trans*-cinnamate product to hippurate and excretes it in the urine. The hippurate acts as a marker for the activity of the bacteria. The bacteria themselves are detectable in feces after dosing, allowing bacterial administration to be regulated. The bacteria are also rendered auxotrophic, thereby providing for their biocontainment. This study

represents a substantial advance in the translational use of bacterial therapeutics to treat metabolic disorders.

Dorival-García N, Carillo S, Ta C, Roberts D, Comstock K, Lofthouse S, Ciceri E, D'Silva K, Kierans G, Kaisermayer C, Lindeberg A, Bones J. Large-scale assessment of extractables and leachables in single-use bags for biomanufacturing. *Analytical Chemistry* 90;2018:9006–9015.

Disposable filter cartridges, disposable tubing and depth filters, disposable columns, and single-use bags are all routinely used in biopharmaceutical manufacturing. From the polymeric materials used to make these devices, compounds may become extracted or leached into the drug product stream. Accordingly, part of manufacturing processes is testing for the presence and safety of extractables and leachables. The present paper contributes to the standardization of this aspect of drug manufacture by providing an inventory of extractable and leachable compounds from 34 films used to make single-use bags. Bags are provided by various suppliers over various dates. Compounds are extracted by methods representing reasonable worst-case conditions. Leachables are found when the drug product contacts the surface under standard operating conditions. Semivolatile compounds are analyzed by non-targeted GC-Orbitrap MS/MS. Nonvolatiles are analyzed by untargeted LC-Orbitrap MS/MS. Trace elements are identified by ICP-MS. The major extractable compounds identified in the study are additives and their degradation products, impurities from the film-manufacturing process, and polymer-related compounds. Among the leachables are compounds not also found among the extractables, indicating that contact with the drug product and processing fluid is indeed required to elicit their appearance. It is hoped that the library of compounds identified in the study will suggest practical guidelines to screen for these substances and to eliminate the harmful ones.